arXiv:0804.1392v2 [stat.CO] 30 Jan 2009

Coverage Probability of Wald Interval for Binomial Parameters *

Xinjia Chen

Submitted in April, 2008

Abstract

In this paper, we develop an exact method for computing the minimum coverage probability of Wald interval for estimation of binomial parameters. Similar approach can be used for other type of confidence intervals.

1 Wald Interval

Let X be a Bernoulli random variable with distribution $\Pr\{X = 1\} = 1 - \Pr\{X = 0\} = p \in (0, 1)$. Let $X_1, \cdots, X_n$ be i.i.d. random samples of X. Let $K = \sum_{i=1}^n X_i$. The widely-used Wald interval is $[L, U]$ with lower confidence limit

$$L = \frac{K}{n} - \mathcal{Z}_{\delta/2} \sqrt{\frac{\frac{K}{n}(1 - \frac{K}{n})}{n}}$$

and upper confidence limit

$$U = \frac{K}{n} + \mathcal{Z}_{\delta/2} \sqrt{\frac{\frac{K}{n}(1 - \frac{K}{n})}{n}}$$

where $\mathcal{Z}_{\delta/2}$ is the critical value such that $\int_{\mathcal{Z}_{\delta/2}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{\delta}{2}$. It has been discovered by Brown et al. [1] that the coverage probability of Wald interval is surprisingly poor.

2 Coverage Probability

The coverage probability of Wald interval for binomial parameters has been investigated by [1] and other researchers by Edgeworth expansion method and numerical methods based on discretizing the binomial parameter. Here, we have obtained expression of the minimum coverage probability of Wald interval, which requires only finite many evaluations of coverage probability.

*The author is currently with Department of Electrical Engineering, Louisiana State University at Baton Rouge, LA 70803, USA, and Department of Electrical Engineering, Southern University and A&M College, Baton Rouge, LA 70813, USA; Email: chenxinjia@gmail.com

Theorem 1 *Define*

$$\mathrm{T}^-(p) = \frac{2p + \theta - \sqrt{\theta^2 + 4\theta p(1-p)}}{2(1+\theta)}, \qquad \mathrm{T}^+(p) = \frac{2p + \theta + \sqrt{\theta^2 + 4\theta p(1-p)}}{2(1+\theta)}$$

with $\theta = \frac{\mathcal{Z}_{\delta/2}^2}{n}$. *Define*

$$\mathscr{L}(k) = \frac{k}{n} - \mathcal{Z}_{\delta/2}\sqrt{\frac{\frac{k}{n}(1-\frac{k}{n})}{n}}, \qquad \mathscr{U}(k) = \frac{k}{n} + \mathcal{Z}_{\delta/2}\sqrt{\frac{\frac{k}{n}(1-\frac{k}{n})}{n}}$$

for $k = 0, 1, \cdots, n$. *Define*

$$C_l(k) = \Pr\{\lceil \mathrm{T}^-(\mathscr{L}(k)) \rceil \leq K \leq k-1 \mid \mathscr{L}(k)\}, \quad C_l'(k) = \Pr\{\lfloor \mathrm{T}^-(\mathscr{L}(k)) \rfloor + 1 \leq K \leq k-1 \mid \mathscr{L}(k)\}$$

for $k \in \{0, 1, \cdots, n\}$ *such that* $0 < \mathscr{L}(k) < 1$. *Define*

$$C_u(k) = \Pr\{k+1 \leq K \leq \lfloor \mathrm{T}^+(\mathscr{U}(k)) \rfloor \mid \mathscr{U}(k)\}, \quad C_u'(k) = \Pr\{k+1 \leq K \leq \lceil \mathrm{T}^+(\mathscr{U}(k)) \rceil - 1 \mid \mathscr{U}(k)\}$$

for $k \in \{0, 1, \cdots, n\}$ *such that* $0 < \mathscr{U}(k) < 1$.

Suppose $\theta < 3$. *Then, the following statements hold true:*

(I): $\inf_{p \in (0,1)} \Pr\{L \leq p \leq U \mid p\}$ *equals to the minimum of*

$$\{C_l(k) : 0 \leq k \leq n;\ 0 < \mathscr{L}(k) < 1\} \cup \{C_u(k) : 0 \leq k \leq n;\ 0 < \mathscr{U}(k) < 1\}.$$

(II): $\min_{p \in (0,1)} \Pr\{L < p < U \mid p\}$ *equals to the minimum of*

$$\{C_l'(k) : 0 \leq k \leq n;\ 0 < \mathscr{L}(k) < 1\} \cup \{C_u'(k) : 0 \leq k \leq n;\ 0 < \mathscr{U}(k) < 1\}.$$

The proof of Theorem 1 is given in the next section.

3 Proof of Theorem 1

We need some preliminary results.

Lemma 1 *For* $n > \frac{\mathcal{Z}_{\delta/2}^2}{3}$, *both the lower and upper confidence limits of Wald interval are monotonically increasing with respect to* k.

Proof. For simplicity of notation, let

$$z = \frac{k}{n}.$$

Then, the upper confidence limit can be written as

$$h(z) = z + \mathcal{Z}_{\delta/2}\sqrt{z(1-z)/n}.$$

Similarly, the lower confidence limit can be written as

$$g(z) = z - \mathcal{Z}_{\delta/2}\sqrt{z(1-z)/n}.$$

To show that the upper confidence limit is monotonically increasing with respect to k, it suffices to show that

$$\frac{\partial h(z)}{\partial z} > 0$$

if $0 \leq h(z) \leq 1$. Since

$$\frac{\partial h(z)}{\partial z} = 1 + \frac{\sqrt{\theta}}{2}\frac{1-2z}{\sqrt{z(1-z)}},$$

which is clearly positive for $0 < z \leq \frac{1}{2}$, it remains to show

$$\sqrt{\theta}\frac{1-2z}{\sqrt{z(1-z)}} > -2, \qquad \forall z \in \left(\frac{1}{2}, 1\right)$$

or equivalently,

$$z(1-z) > \frac{\theta}{4}(2z-1)^2, \qquad \forall z \in \left(\frac{1}{2}, 1\right).$$

Note that $h(z) < 1$ for $0 < z < \frac{1}{1+\theta}$, and $h(z) > 1$ for $1 > z > \frac{1}{1+\theta}$.

If $\frac{1}{1+\theta} \leq \frac{1}{2}$, i.e., $\theta \geq 1$, then we are done, since $\frac{\partial h(z)}{\partial z} > 0$ for all $0 < z < \frac{1}{1+\theta} \leq \frac{1}{2}$. Otherwise, if $\theta < 1$, it suffices to show

$$w(z) = z(1-z) - \frac{\theta}{4}(2z-1)^2 > 0$$

for $\frac{1}{2} < z < \frac{1}{1+\theta}$. Since $z(1-z)$ is decreasing and $\frac{\theta}{4}(2z-1)^2$ is increasing for $1 > z > \frac{1}{2}$, $w(z)$ is decreasing for $\frac{1}{2} < z < \frac{1}{1+\theta}$. Therefore, it suffices to have

$$w\left(\frac{1}{1+\theta}\right) > 0,$$

i.e.,

$$\frac{1}{1+\theta}\left(1 - \frac{1}{1+\theta}\right) - \frac{\theta}{4}\left(\frac{2}{1+\theta} - 1\right)^2 > 0,$$

i.e.,

$$\frac{\theta}{(1+\theta)^2} - \frac{\theta}{4}\left(\frac{1-\theta}{1+\theta}\right)^2 > 0,$$

i.e.,

$$4 > (1-\theta)^2,$$

which is guaranteed since $0 < \theta < 3$. This shows that the upper confidence limit is monotonically increasing with respect to k. Observing that $g(z) = 1 - h(1-z)$, we have that the lower confidence limit is also monotonically increasing with respect to k.

$\square$

Now we consider the minimum coverage probability. By solving equation

$$\left(p - \frac{k}{n}\right)^2 = \theta \frac{k}{n}\left(1 - \frac{k}{n}\right)$$

with respect to k, we can show that

$$\Pr\{L \leq p < U \mid \mathscr{U}(k)\} = \Pr\{k < K \leq \mathrm{T}^+(p) \mid \mathscr{U}(k)\}, \qquad 0 < \mathscr{U}(k) < 1$$

$$\Pr\{L < p \leq U \mid \mathscr{L}(k)\} = \Pr\{\mathrm{T}^-(p) \leq K < k \mid \mathscr{L}(k)\}, \qquad 0 < \mathscr{L}(k) < 1$$

$$\Pr\{L < p < U \mid \mathscr{U}(k)\} = \Pr\{k < K < \mathrm{T}^+(p) \mid \mathscr{U}(k)\}, \qquad 0 < \mathscr{U}(k) < 1$$

$$\Pr\{L < p < U \mid \mathscr{L}(k)\} = \Pr\{\mathrm{T}^-(p) < K < k \mid \mathscr{L}(k)\}, \qquad 0 < \mathscr{L}(k) < 1$$

Since both the lower and upper confidence limits of Wald interval are monotone as asserted by Lemma 1, the proof of Theorem 1 can be completed by making use of the above results and applying the theory of coverage probability of random intervals established by Chen in [2].

References

[1] Brown, L. D., Cai, T. and DasGupta, A., "Interval estimation for a binomial proportion and asymptotic expansions," *The Annals of Statistics*," Vol. 30, pp. 160-201, 2002.

[2] Chen X., "Coverage Probability of Random Intervals," arXiv:0707.2814, July 2007.